2016-08-25 Meeting notes

Date

☷ 25 Aug 2016

Attendees

- Kathleen Carr
- Alicia Aleman
- Kaylin Bugbee
- Nathan James
- Jeanne' le Roux
- Tyler Stevens

Goals

- Discuss timeline

    Dec/Jan: DAACs receive results of automated unified curation scripts

    Summer 2017: DAACs correct errors flagged in automated reports

    TBD: Manual curation reports to DAACs

- Pass #2 through the test record or through the generic CMR record

Discussion Items

Item	Who	Notes
Timeline	all	Kathy presented the planned timeline from Andy Mitchell, i.e., Dec/Jan: DAACs receive results of automated unified curation scripts Summer 2017: DAACs correct errors flagged in automated reports TBD: Manual curation reports to DAACs
Unified curation tool - what form will it take?	all	GCMD uses a UI to curate metadata ARC uses a script to curate metadata We need to determine whether the unified curation tool will be a UI or a script or something else We also need to determine the form and content for the output from the unified curation tool - both GCMD and ARC currently write curation reports to a spreadsheet
Unified curation rules	all	Kathy is accumulating curation rules from our past two sessions in a spreadsheet which has columns for DIF9, DIF 10, ECHO 10, and UMM-C fields for Collection records We looked at the spreadsheet together during the meeting. Kathy will send it out for the group to review and comment.

Unified curation rules for different native metadata formats	all	How does automated curation script interact with the native format of the record – e.g. DIF, DIF 10, ECHO 10, UMM-JSON, ISO Answer: there is a Rule set for each native format Discussion: 1. Can we make an assumption that each DAAC has a single native metadata format, based on the ECHO vs DIF choices made last year during the Reconciliation effort? One counter example is CDDIS, who has recently added and updated collection records via the MMT, so for those records, the native format is now UMM-JSON. 2. Should we create a different curation tool for each native format, or create a single curation tool which first determines the native format of the record and then branches to the rules for that format?
When metadata errors are reported during curation, what recommendations do we make to the DAACs regarding how to fix the errors?	all	For errors related to invalid values for controlled vocabulary, invalid value formats, and invalid value ranges, the curation report produced by automated curation should include a reference to the controlled vocabulary source (e.g., KMS, enum list), the valid format, and the valid value range respectively. Follow-up with specific recommendations from the curation team would be part of the manual curation process.
Questions about fields in the KMS not in the metadata schemas	all	•Platform – should metadata contain the Series Entity which is in the KMS for Platforms? No. Therefore, no schema updates or additional curation checks needed. (maybe a new use case for EDSC Faceted search?) •Instrument – should metadata contain Category,Class,Type,Subtype? No. Therefore, no schema updates or additional curation checks needed (maybe a new use case for EDSC Faceted search) •Do we want to make a general recommendation that case in metadata should match case in KMS for all KMS keywords? Yes; if not flag error
Potential KMS updates that we have been noting	all	•Temporal keywords (also see MMT beta test comments and related ticket opened a few months ago by Tyler) •Platform long name for CMORPH •Computer as instrument short name and long name •Compare KMS list of URL Content types with UMM-C list (see Related URL in the current UMM-C review; Simon Cantrell is author, and Scott Ritz has entered comments)
Metadata correction issues where a working group may need to determine metadata policy (so that the DAACs know how to correct the issue)	all	•Do we need additional and / or different values in the KMS for Temporal Keywords? •How to enter multiple instances of Data Format (separate instances, single instance with comma-separated values) •Mixed case for KMS keywords? (consider work in progress to allow options on API call to choose case) •GHRC field campaign data vis a vis Location Keywords •How to handle multiple spatial extents in a single collection

Action items

- [ ] Kathleen Carr Talk to Dana, then set up a meeting with the curation team plus GCMD and ARC developers to begin the process of populating the ECC Curation repository with unified curation tool code
- [x] Kathleen Carr Send curation rules spreadsheet to group for review and comment